Learning from Candidate Labeling Sets

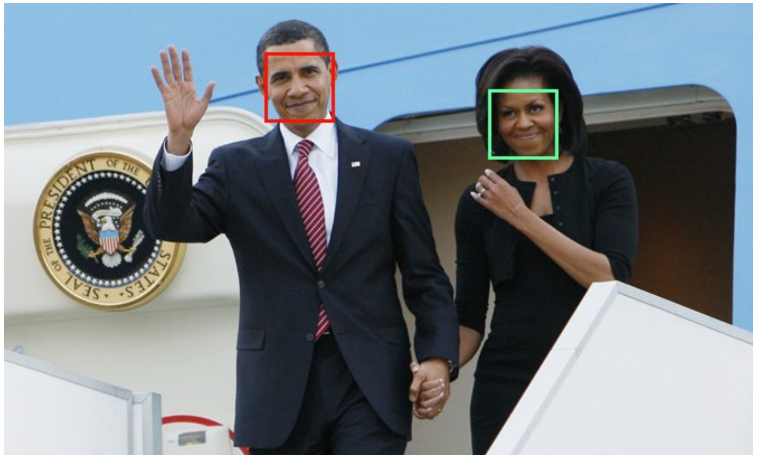
Luo Jie and Francesco Orabona

POSTER M85

- A semi-supervised learning framework that descends from Multiple Label Learning and Multiple Instance Learning
- Encoding prior-information in a principle way
- An ambiguous loss function and a large-margin discrimination formulation
- Efficient algorithm based on stochastic sub-gradient descend

Applications

- Learning from weakly supervised data (Images and Captions)



President ***Barack Obama*** (B) *and first lady* ***Michelle Obama*** (M) *wave from the steps of Air Force One as they arrive in Prague, Czech Republic.*

	$Face_{left}$	$Face_{right}$
$Z_{1:}$	B	M
$Z_{2:}$	M	B
$Z_{3:}$	B	null
$Z_{4:}$	null	B
$Z_{5:}$	M	null
$Z_{6:}$	null	M

null: not shown in the caption

- Learning from multiple annotators

- Clustering with constraints

We introduce an ambiguous loss which upper bounds the true loss in expectation (by a proportionality factor)

$$\ell_{\max}\left(\mathcal{X}, \mathcal{Z}; \boldsymbol{w}\right) = \left| \max_{\bar{\boldsymbol{z}} \notin \mathcal{Z}} \left(\ell^{A}_{\Delta}(\bar{\boldsymbol{z}}, \mathcal{Z}) + \mathbf{F}(\mathcal{X}, \bar{\boldsymbol{z}}; \boldsymbol{w})\right) - \max_{\boldsymbol{z} \in \mathcal{Z}} \mathbf{F}(\mathcal{X}, \boldsymbol{z}; \boldsymbol{w}) \right|_{+}$$

$$\text{where} \quad \ell^{A}_{\Delta}(\boldsymbol{z}, \mathcal{Z}) = \min_{\boldsymbol{z}' \in \mathcal{Z}} \ell_{\Delta}(\boldsymbol{z}, \boldsymbol{z}')$$

Non-Convex

Every instance x_m from a bag $\mathcal{X}$ is considered jointly at prediction, the prior knowledge between instances is encoded in the labeling vector y

$$\mathbf{F}(\mathcal{X}, \boldsymbol{y}; \boldsymbol{w}) = \sum_{m=1}^{M} F(\boldsymbol{x}_m, y_m) = \sum_{m=1}^{M} \boldsymbol{w} \cdot \phi(\boldsymbol{x}_m) \otimes \psi(y_m)$$

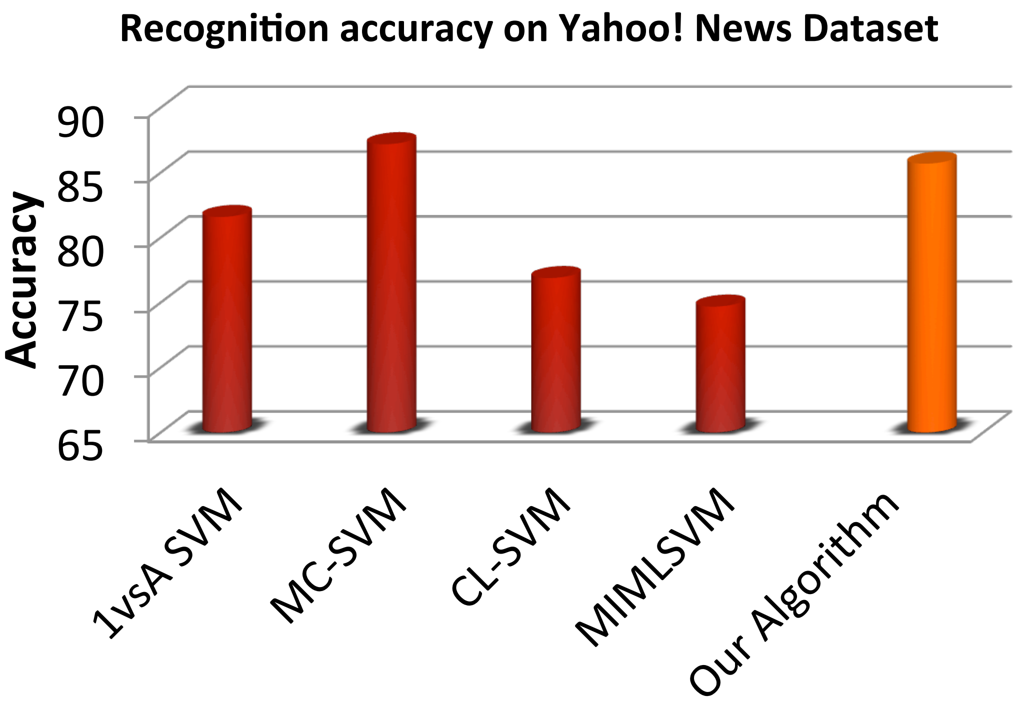
We propose a large margin optimization problem based on the ambiguous loss and solve the problem using the Constrained Concave-Convex Procedure (CCCP)

Solve the CCCP-relaxed optimization problem using a stochastic sub-gradient descent algorithm based on the Pegasos framework

Poster Loaction: M85

Code Available

Yahoo! News dataset [Berg, et.al. NIPS04]



- Performance comparable to fully-supervised SVM algorithms (the 1st and 2nd columns)
- Outperform other weakly-labeled learning baselines (the 3rd and 4th columns)